MINING STUDENTS' PERFORMANCE IN UPSR USING STATISTICS AND NEURAL NETWORKS

NOR FAZIDA ABD. RAHMAN

UNIVERSITI UTARA MALAYSIA
2008

MINING STUDENT'S PERFORMANCE IN UPSR USING STATISTICS AND NEURAL NETWORKS

A thesis submitted to the Division of Applied Sciences, College of Arts and Sciences in partial fulfillment of the requirements for the degree Master of Science (Information and Communication Technology), Universiti Utara Malaysia.

by
Nor Fazida Abd Rahman

KOLEJ SASTERA DAN SAINS
(College of Arts and Sciences)
Universiti Utara Malaysia

PERAKUAN KERJA KERTAS PROJEK
(Certificate of Project Paper)

Saya, yang bertandatangan, memperakukan bahawa
(I, the undersigned, certify that)

NOR FAZIDA ABD. RAHMAN

calon untuk Ijazah
(candidate for the degree of) **MSc. (ICT)**

telah mengemukakan kertas projek yang bertajuk
(has presented his/her project paper of the following title)

MINING STUDENTS PERFORMANCE IN UPSR USING STATISTICS AND NEURAL NETWORKS

seperti yang tercatat di muka surat tajuk dan kulit kertas projek
(as it appears on the title page and front cover of project paper)

bahawa kertas projek tersebut boleh diterima dari segi bentuk serta kandungan dan meliputi bidang ilmu dengan memuaskan.
(that the project paper acceptable in form and content, and that a satisfactory knowledge of the field is covered by the project paper).

Nama Penyelia Utama
(Name of Main Supervisor): **ASSOC. PROF. FADZILAH SIRAJ**

Tandatangan
(Signature) : Fadzilah

Tarikh
(Date) : 28/5/2008

PERMISSION TO USE

In presenting this thesis in partial fulfillment of the requirements for a postgraduate degree from Universiti Utara Malaysia, I agree that the University Library may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purpose may be granted by my supervisor(s) or, in their absence by the Dean of College of Arts and Sciences. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to Universiti Utara Malaysia for any scholarly use which may be made of any material from my thesis.

Requests for permission to copy or to make other use of materials in this thesis, in whole or in part, should be addressed to

Division of Applied Sciences
College of Arts and Sciences
Universiti Utara Malaysia
06010 UUM Sintok
Kedah Darul Aman.

ABSTRACT (BAHASA MELAYU)

Pencapaian pelajar di dalam peperiksaan menjadi penanda aras yang penting dalam menentukan kualiti pendidikan di Malaysia. Data-data peperiksaan telah dikumpul mulai ujian-ujian bulanan yang telah dijalankan sehingga ke percubaan UPSR untuk diuji dengan peperiksaan UPSR yang sebenar. Ini juga melibatkan data-data lain yang berkaitan seperti latar belakang keluarga dan maklumat berkenaan persekolahan pelajar. Data mentah diproses serta dianalisa menggunakan kaedah Statistik. Kaedah Statistik memberikan analisis yang bernilai kepada model pencapaian. Kemudian, kombinasi unit input, unit tersembunyi dan unit output diuji untuk meramal pencapaian sebenar pelajar. 5 model diuji berdasarkan 5 matapelajaran teras untuk mengaitkannya dengan faktor-faktor lain menggunakan analisis diskriptif. Justeru, hubungan itu dikaji dengan teliti untuk mengukuhkan model jangkaan. Model akhir kes ini menggunakan kadar pembelajaran 0.1, kadar momentum 0.1, fungsi aktivasi Sigmoid, kriteria penghentian pembelajaran 100 pusingan dengan senibinanya 13 unit input, 2 unit tersembunyi dan 5 unit output. Keputusan yang telah diperolehi menunjukkan Rangkaian Neural mempunyai potensi yang tinggi untuk meramal pencapaian pelajar di masa hadapan.

ABSTRACT (ENGLISH)

Academic performance has become an important evidence of determining the quality in Malaysia's education system. The examination data is collected on the previous students' examinations yet to be tested for their coming UPSR. The other related data such as family background and schooling information are also involved. The raw data is preprocessed and analyzed using statistical method. The results from the statistical analysis indicate the significant contribution of these attributes to the achievement model. The combinations of input variables, hidden layer and output nodes are explored to predict the students' performance. Five models are constructed based on five subjects to relate them with other factors for the purpose of descriptive analysis. The relationship between examination results and other factors are investigated thoroughly to enhance the prediction model. The performance model obtained in this study uses parameters such as; learning rate 0.1, momentum rate 0.1, Sigmoid activation function, 100 epoch learning stopping criteria with its architecture, 13 inputs unit, 2 hidden units and 5 output units. The result indicates that Neural Networks has high potential to be used in predicting students' performance.

ACKNOWLEDGMENT

A million of thanks are due first to my supervisor, Associate Professor Fadzilah Siraj for patiently navigating and generously sharing her rich source of knowledge with me. A good teacher always guides her students in no matter situation they are.

Equally thankful to my family for all the love and support by giving me a beautiful place in their heart. Fortunately born in teachers' family, education is the top priority to me. In a process towards seeking for excellent and meaningful education, knowledge brings its own values. Heartfull of thanks also flow to my sister, Ihwana As'ad for her loving and caring towards me.

A warm thanks to my colleagues; SK Tan Sri Ghazali Jawi, SK Basia Lama, SK Sri Adika Raja and SK Mahkota Sari, Gerik, Perak for allowing me to use their schooling dataset as experimental tool. These research findings are specially dedicated for the sake of my sacred 'children'.

I would like to express my eternal gratitude to my dearest partner, Mohd Shukri Mohd Adam for his kindness and for everlasting bond of fellowship. Not least, my deep appreciation goes to all who lend a hand in realizing this project.

Nor Fazida Abd Rahman

Division of Applied Sciences

College of Arts and Sciences

2008

TABLE OF CONTENTS

CHAPTER 3: METHODOLOGY

CHAPTER 4: RESULTS AND DISCUSSION

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AI	Artificial Intelligence
ANN	Artificial Neural Networks
BI	Bahasa Inggeris
BM1	Bahasa Melayu 1
BM2	Bahasa Melayu 2
BP	Backpropagation Algorithm
FF	Feed Forward Algorithm
MLP	Multilayer Perceptrons
MT	Mathematics
NC	Neural Connection
NN	Neural Networks
SAPR 16	Sistem Analisis Peperiksaan
SC	Science
SK	Sekolah Kebangsaan
SMM	Sistem Maklumat Murid
PKSR	Penilaian Kemajuan Berasaskan Sekolah
UPSR	Ujian Penilaian Sekolah Rendah

CHAPTER 1

INTRODUCTION

This section discusses the background of the study that consists of general overview on the education system in Malaysia, some brief description on the problem stastements, objectives, scope and significance of this study. Finally, this section presents the thesis organization describing the structure of this report.

1.1 Overview

Realizing the importance of having a national database that comprises of graduate data, the Ministry of Higher Institutions of Learning has taken an initiative to form such a data warehouse. At the school level, 'Sistem Maklumat Murid' (SMM) is used for several years to collect information on students' background. SMM consists mainly information about demographic and co-curricular but not including their academic results. To support SMM, 'Analisis Peperiksaan 16' (SAPR 16) is used for the purpose of analyzing the students' results and ranks their performance accordingly. Even then, the analysis involves the use of descriptive approach of data mining technique. Little attempt has been made to include the use of forecasting model in managing the students'

The contents of the thesis is for internal user only

REFERENCES

Larose D., (2006). *Data Mining Methods and Models.* New Jersey: John Wiley & Sons Inc, 93-149.

Hand D., Mannila H., Smyth P., (2001). *Principles of Data Mining.* Cambridge: MIT Press.

Haykin S., (2007). *Neural Networks: A Comprehensive Foundation.* New York: Macmillan Publishing, 156-175.

Khoa N., Sakakibara K., Nishikawa I., (2006). Stock Price Forecasting Using Back Propagation Neural Networks with Time and Profit Based Adjusted Weight Factors. *Electronics & Communication Engineering Journal,* 5484 – 5488.

Fung C., Iyer V., Brown W., Wong K., (2005). Comparing The Performance of Different Neural Networks Architectures for The Prediction of Mineral Prospectivity. *Electronics & Communication Engineering Journal,* vol. 1, 394 – 398.

Wang T., Mitrovic A., (2002). Using Neural Networks to Predict Student's Performance. *Electronics & Communication Engineering Journal,*969 – 973.

Huisken G., Coffa A., (2000). Short-term Congestion Prediction: Comparing Time Series with Neural Networks. *Electronics & Communication Engineering Journal,* 66 – 69.

Cripps A., (1996). Using Artificial Neural Nets to Predict Academic Performance. *American Psychological Association Journal*, 33 – 37.

Hostetler T., (1996). Predicting Student Success in An Introductory Programming Course. *American Psychological Association Journal,* pp. 40 – 49.

Butcher D., Muth W., (1985). Predicting Performance in An Introductory Computer Science Course. *American Psychological Association Journal* , 263 – 268.

Bae, Yupin, Smith, Thomas M., (1997). Women in Mathematics and Science. *Education Resources Information Center Journal.*

Gibson, Margaret A., (2003). Improving Graduation Outcomes for Migrant Students. *Education Resources Information Center Journal.*

Terry E., Spradlin, Kirk R., Walcott C., Kloosterman P., Zaman K., McNabb S., Zapf J. & associates, (2005). Is The Achievement Gap in Indiana Narrowing. *Education Resources Information Center Journal.*

Hayek, John C., Kuh, George D., (1999). College Activities and Environmental Factors Associated with The Development of Life Long Learning Competencies of College Seniors. *Education Resources Information Center Journal.*

Erbe, Mach B., (2000). Correlates of School Achievement in Chicago Elementary Schools. *Education Resources Information Center Journal.*

Henchey, Norman, (2002). Schools That Make A Difference : Final Report. Twelve Canadian Secondary Schools in Low Income Settings. *Education Resources Information Center Journal.*

Dai, Huang, (2006). Data Mining Used in Rule Design for Active Database Systems. *Electronics & Communication Engineering Journal.*

Young, Ho, Il., Ok, Suk, (2003). The Reliability of Pollution with Regression Analysis and The Possibility of Dispersion and Receptor Models. *Electronics & Communication Engineering Journal.*

Yusof & Syed Hassan, (2000). Regression Analysis of Consonant Frequency over Number of Turn of Normal Mode Helical Antenna. *Electronics & Communication Engineering Journal.*

Bon, Ogier, Razali, (2007). Modelling in Manufacturing Industry : Parameters Selection Using Regression Analysis. *Electronics & Communication Engineering Journal.*

Perez, Gonzalez, Salinas, (2000). Neural versus Difference Equation Modeling for 2D Pattern Recognition Problems. *Electronics & Communication Engineering Journal.*

Fausett, Elwasif, (1994). Predicting Performance From Test Scores Using Backpropagation and Counterpropagation . *Electronics & Communication Engineering Journal.*

Burtner, (2004). Critical-To-Quality Factors Associated with Engineering Student Persistence: The Influence of Freshman Attitudes. *Electronics & Communication Engineering Journal.*

Han, Cheng, Meng, (2003). Application of Four-layer Neural Network on Information Extraction. *Electronics & Communication Engineering Journal.*

Walczak, (1994). Categorizing University Student Applicants with Neural Networks. *Electronics & Communication Engineering Journal.*

Mahmoud E., (2004). Accuracy in Forecasting: A Survey. *Journal of Forecasting.*

Friedman D., Montgomery D., (1985). Evaluation of The Predictive Performance of Biased Regression Estimators. *Journal of Forecasting.*

Tsukamoto Y., Namatame A., (1996) Evolving Neural Network Models. *Electronics & Communication Engineering Journal.*

Phua C., Alahakoon D., Lee V. (2000). Minority Report in Fraud Detection: Classification of Skewed Data. *Sigkdd Explorations.*

Schultz R., (1984). The Implementation of Forecasting Models. *Journal of Forecasting.*

Klein L., (1984). The Importance of The Forecast. *Journal of Forecasting.*

Huss W., (1985). The Teachers/Practitioners Corner Comparative Analysis of Load Forecasting Techniques at a Southern Utility. *Journal of Forecasting.*

Gruca T., Klemz B., Petersen E., (1999). Mining Sales Data Using A Neural Network Model of Market Response. *SIGKDD Explorations.*

Lu C., Brabanter J., Huffel S., Vergote I., Timmeman D. (2001). Using Artificial Neural Networks To Predict Malignancy of Ovarian Tumors. *Proceedings of the 23rd Annual EMBS International Conference.*

Atiya A., El-Shoura S., Shaheen S., El-Sherif M. (1999). "A Comparison Between Neural-Network Forecasting Techniques – Case Study: River Flow Forecasting" *Electronics & Communication Engineering Journal.*

Ringwood J., Galvin G. (2002). Computer-Aided Learning in Artificial Neural Networks. *Electronics & Communication Engineering Journal.*

Burke H., Rosen D., Goodman P., (1994). Comparing Artificial Neural Networks To The Other Statistical Methods For Medical Outcome Prediction. *Electronics & Communication Engineering Journal.*

Sitte R., Sitte J. (2000). Analysis Of The Predictive Ability Of Time Delay Neural Networks Applied To The S&P 500 Time Series. *Electronics & Communication Engineering Journal.*

Jiuzhen L., Jiaqing Z., (2002). Procedure Neural Networks With Supervised Learning. *Proceeding of the 9th International Conference on Neural Information Processing.*

Ritschel, Pfeifer W., Grob (1994). Rating of Pattern Classifications in Multi-layer Perceptrons: Theoretical Background And Practical Results. *ACM.*

Mahamud K., Bakar A., Norwawi N. (1998). Neural Network Modelling To Predict House Prices Performance. *Research Report.*

Roy A., (2000). Artificial Neural Networks – A Science in Trouble. *SIGKDD Explorations,* 33.

Chamillard A., (2006). Using Student Performance Predictions in A Computer Science Curriculum. *ACM.*

Spradlin T., Kirk R., Walcott C., Kloosterman P., Zaman K., McNabb S., Zapf J., (2005). Is The Achievement Gap in Indiana Narrowing? *Center for Evaluation & Education Policy.*

Sweet J., Rasher S., Abromitis B., Johnson E., (2004). Case Studies of High-Performing, High-Technology Schools. *North Central Regional Educational Laboratory.*

Williams, Bugg D., (1998). Parent Involvement Gender Effects On Preadolescent Student Performance. *American Educational Research Association.*

Verna, Ann M., Spina, Maria, (2002). Parental Processes and Self-Concepts Effect The Academic Achievement of Italian Students. *American Educational Research Association.*

Sengupta S., (1995). A Comparative Study of Neural Network and Regression Analysis As Modelling Tools. *Electronics & Communication Engineering Journal.*

Li Y., (1994). A General Linear-Regression Analysis Applied To The 3-Parameter Weibull Distribution. *Electronics & Communication Engineering Journal.*

Itsuki R., Yajima H., Mizuno H., Kinukawa H. (1996). Application and Verification of Using Statistical Analysis Tool and Expert System Together in Multiple Regression Analysis. *Electronics & Communication Engineering Journal.*

Fukuda T., Shibata T., Tokita M., Mitsuoka T. Neural Network Application for Robotic Motion Control. *Research Report*

Wong K., (1993). Artificial Intelligence and Neural Network Applications In Power Systems. *Electronics & Communication Engineering Journal.*

Means R., (1994). High Speed Parallel Hardware Performance Issues for Neural Network Applications. *Electronics & Communication Engineering Journal.*

El-Sharkawi M., Huang S., (1994). Ancillary Techniques for Neural Network Applications. *Electronics & Communication Engineering Journal.*

Vonk E., Jain L., Veelenturf L., (1995). Neural Network Applications. *Electronics & Communication Engineering Journal.*

Sugiyama S, A Neural Network Application To Semantic and Logistic Recognition. *Electronics & Communication Engineering Journal,* 1995.

Bershad N., Ibnkahla M., Castanie F., (1997). Statistical Analysis of A Two-Layer Backpropagation Algorithm Used For Modelling Nonlinear Memoryless Channels: The Sigle Neuron Case. *Electronics & Communication Engineering Journal.*

Hansen J., Nelson R., (1997). Neural Networks and Traditional Time Series Methods: A Synergistic Combination in State Economic Forecasts. *Electronics & Communication Engineering Journal.*

Ye Q., Liang B., Li Y., (2005). Amnestic Neural Network for Classification: Application on Stock Trend Prediction. *Electronics & Communication Engineering Journal.*

Lisboa P., Taylor M., (1993). Workshop on Neural Network Applications and Tools. *Electronics & Communication Engineering Journal.*

Tal B., (2003). Background Information On Our Neural Network – Based System of Leading Indicators. *CIBC World Markets.*

Bojkovic Z., Milovanovic D., Mastorakis N., (2000). Neural Networks Applications For Multimedia Processing. *Electronics & Communication Engineering Journal.*

Chowdhury F., Wahi P., Raina R., Kaminedi S., (2001). A Survey of Neural Networks Applications in Automatic Control. *Electronics & Communication Engineering Journal.*

Rajagopalan R., Rajagopalan P., (1996). Applications of Neural Network in Manufacturing. *Electronics & Communication Engineering Journal.*

Bose B., (2007). Neural Network Applications in Power Electronics and Motor Drives – An Introduction and Perspective. *Electronics & Communication Engineering Journal.*

Wikipedia, (2008). Statistics. http://.en.wikipedia.org/wiki/Statistics.